High level design (HLD)

Thyroid Disease Detection

Table of Contents

**Abstract2**

**1 Introduction3**

1.1 Why this High-Level Design Document3

1.2 Scope3

**2 General Description**4

2.1 Product Perspective4

2.2 Problem Statement4

2.3 Proposed Solution4

2.4 Further Improvements4

2.5 Technical Requirements5

2.6 Data Requirements5

2.7 Tools Used5

2.8 Constraints6

2.9 Assumptions6

**3 Design Details**7

3.1 Process Flow7

3.1.1 Proposed Methodology7

3.1.2 Model Training and Evaluation7

3.1.3 Deployment Process7

3.2 Event Log8

3.3 Performance8

3.4 Reusability8

3.5 Application Compatibility8

3.6 Resource Utilization9

3.7 Deployment9

**4 Conclusion**10

**Abstract**

Thyroid disease a very common problem in India especially in female. Thyroid disorder can speed up or slow down the metabolism of the body. Hyperthyroidism and hypothyroidism are the most two common diseases caused by irregular functioning of thyroid gland. Machine learning plays an important role in predicting the disease. In the present work different machine learning algorithms like logistic algorithm, Decision Trees, Random Forest, Support Vector Machine and XG Boost algorithm are used for classification of thyroid disease. Based the obtained result Decision Trees algorithm is used to predict the thyroid disease. Web app is created to get data from users to predict the type of disease using flask. The user input and result is stored in Cassandra cloud database.

**1 Introduction**

**1.1 Why this High-level Design document?**

High-level design (HLD) explains the architecture that would be used to develop a system. The architecture diagram provides an overview of an entire system, identifying the main components that would be developed for the product and their interfaces. The HLD uses possibly nontechnical to mildly technical terms that should be understandable to the administrators of the system.

The HLD will -

* Present all of the design aspects and define them in detail.
* Describe the user interface being implemented.
* Describe the hardware and software interfaces.
* Describe the performance requirements.
* Include design features and the architecture of the project.
* List and describe the non-functional attributes like:
  + Security
  + Reliability
  + Maintainability
  + Portability
  + Reusability
  + Application compatibility
  + Resource utilization
  + Serviceability

**1.2 Scope**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. It uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

**2 General Description**

**2.1 Product perspective:**

The Thyroid disease detection is a machine learning based multi classification model which will help us to detect the thyroid disease whether it is hyperthyroid or hypothyroid thyroid and take the necessary action. It uses seven features like age, sex, TSH, T3, FTI, TT4 and T4U to predict the risk of thyroidal disease.

**2.2 Problem Statement:**

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are one of the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of body's metabolism. The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.

**2.3 Proposed Solution:**

The proposed solution for this project is Machine learning algorithms can be implemented to predict the risk of thyroid disease with 7 features mentioned above. With those features as inputs from the web app, the implemented classification model will predict the output as hyperthyroid, hyperthyroid or absence of the thyroid. Finally, based on the given prediction the medical team can take necessary steps

**2.4 Further Improvements:**

The solution can be improved by adding more data from the other sources like hospitals. Also, further feature engineering could result in a better accuracy of our models.

**2.5 Technical Requirements:**

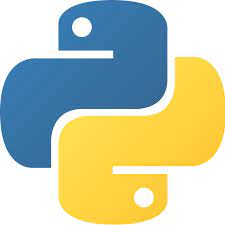
The primary requirement is a device through which our solution can be accessed. Our solution is deployed on Amazon web service (AWS). Thus, one will also require the web address as well.

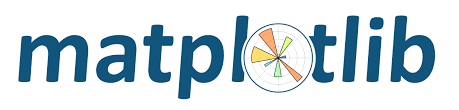
**2.6 Data Requirements:**

The allbp data is used for this project which is taken from the UCI Machine Learning Repository. The given dataset contains 2800 records and 30 features. Features are distributed as 7 continuous features and 23 categorical features. But to run our solution, one needs only 7 features like age, sex, TSH, T3, FTI, TT4 and T4U to predict the risk of thyroidal disease

**2.7 Tools Used:**

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask are used to build the model as shown below.

** **

** **  ** ** 

The description of the tools used is given below

* Jupyter notebook and PyCharm are used as IDE.
* Python is used as our programming language.
* HTML is used to design our frontend.
* Azure is used for deployment of the model.
* Pandas, NumPy and Scikit-learn are used for data processing.
* Flask API is used to integrate our backend with our frontend.
* GitHub is used to store our code.
* Cassandra is used to retrieve, insert, delete, and update the database
* Matplotlib and seaborn are used to visualize our data.

**2.8 Constraints:**

The Thyroid disease detection solution should be user friendly, as automated as possible and users should not be required to know any of the workings.

**2.9 Assumptions:**

Machine learning based solution is used for detecting thyroid disease so it is assumed that there is a linear correlation between the 7 features. Also, correct values of the hormones is expected in order to get a valid prediction.

**3 Design Details**

**3.1 Process Flow**

For identifying thyroidal diseases, we will use a machine learning based model. Below is the process flow diagram.

**3.1.1 Proposed Methodology**

Proposed methodology

**Get data from web App**

**Machine learning model for thyroid detection**

**Display the predicted result in web app**

**Store the user input and result in database**

**3.1.1 Model training and Evaluation**

**Get collection from UCI**

**Data Exploration**

**Data Cleaning**

**Feature Engineering**

**Train test split**

**Model**

**Training set**

**Result of evaluation**

**Testing set**

**3.1.2 Deployment process**

**Start**

**Load model**

**Make prediction**

**Predicted result**

**3.2 Event Log**

The system should log every event so that the user will know what process is running internally.

Initial step by step description:

* The system identifies at what step logging is required
* The system should be able to log each and every system flow
* file logging is used

**3.3 Error Handling**

An error will be defined as anything that falls outside the normal and intended usage. Exception handling is used in the project

**3.3 Performance**

Thyroid disease detection solution is used for detection of thyroid disease in patience based on the amount of hormones present in blood. This will help the medical team in diagnosing thyroidal disease in patience by minimising human error and increasing efficiency. Also, model retraining is very important to improve the performance. The accuracy of the model is 99.30% with decision tree.

**3.4 Reusability**

The code written and the components have the ability to be reused with no problems.

**3.5 Application Compatibility**

The different components for this project will be using Python as an interface between them. Each component will have its own risk task to perform, and it is the job of the Python to ensure proper transfer of information.

**3.6 Resource Utilization**

When any task is performed, it will likely use all the processing power available until that function is finished.

**3.7 Deployment**

The azure cloud platform is used for this entire solution hosting.

**4 Conclusion**

In this project, about four machine learning classification models are evaluated for the given dataset to predict the thyroid disease. The allbp data was used to develop the four classification models, and the predicted the output from these models by comparing with actual target to check the accuracies of these models. It has been found decision tree provides better accuracy compared to others. This solution eliminates human error and increases efficiency in diagnosing thyroid disease.

**5 References**

1. <https://archive.ics.uci.edu/>
2. Google.com for images